

Competing activation mechanisms in epidemics on networks

Claudio Castellano^{1,2} and Romualdo Pastor-Satorras^{3,*}

January 26, 2013

¹Istituto dei Sistemi Complessi (ISC-CNR), Via dei Taurini 19, I-00185 Roma, Italy

²Dipartimento di Fisica, “Sapienza” Università di Roma, P.le A. Moro 2, I-00185 Roma, Italy

³Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain

* Corresponding author

Abstract

In contrast to previous common wisdom that epidemic activity in heterogeneous networks is dominated by the hubs with the largest number of connections, recent research has pointed out the role that the innermost, dense core of the network plays in sustaining epidemic processes. Here we show that the mechanism responsible of spreading depends on the nature of the process. Epidemics with a transient state are boosted by the innermost core. Contrarily, epidemics allowing a steady state present a dual scenario, where either the hub independently sustains activity and propagates it to the rest of the system, or, alternatively, the innermost network core collectively turns into the active state, maintaining it globally. In uncorrelated networks the former mechanism dominates if the degree distribution decays with an exponent larger than $5/2$, and the latter otherwise. Topological correlations, rife in real networks, may perturb this picture, mixing the role of both mechanisms.

The discernment of the mechanisms that contrive to activate spreading processes on heterogeneous substrates is a pivotal issue, with practical applications ranging from the containment of epidemic outbreaks¹ to the viral spreading of rumors and beliefs^{2,3}. The interest on the effects of heterogeneity has been brought about by the observation that social contact networks (the natural substrate for most human epidemic processes) are generally strongly heterogeneous^{4,5,6}, observation that has led to the introduction of complex network theory in the quantitative analysis of epidemic spreading⁷. In this context, the nature of the activation mechanisms translates on simple epidemic models⁸ in setting the epidemic threshold λ_c for some rate of infection λ (the spreading rate), separating a phase in which the spreading affects a finite fraction of the population from a state in which only a vanishingly small fraction is hit. The research effort is thus focused on a twofold objective: The identification of the activation mechanisms as a function of the network topology, and the determination of the functional form of the epidemic threshold.

For the sake of concreteness, we focus our discussion on the simplest models of disease spreading, namely the susceptible-infected-susceptible (SIS) and the susceptible-infected-recovered (SIR) models, leading, respectively, to a steady endemic state or to transient outbreaks affecting a given fraction of the population⁸ (see Methods). On a network substrate—statistically described, at the simplest level, by its degree distribution $P(q)$, defined as the probability that a randomly chosen individual (vertex) is connected to q other individuals⁴—the application of a heterogeneous mean-field (HMF) approach⁹ assuming no topological correlations¹⁰ and neglecting dynamical correlations, yields epidemic thresholds inversely proportional to the second moment of the degree distribution, $\langle q^2 \rangle$ ^{11,12,13}. Since most natural networks have a degree distribution scaling as⁴ $P(q) \sim q^{-\gamma}$, $\langle q^2 \rangle$ takes the form, in the continuous degree approximation, $\langle q \rangle \sim q_{max}^{3-\gamma}$, where q_{max} is the maximum degree in the network⁵. The second moment therefore diverges in the infinite network size limit (i.e. when $q_{max} \rightarrow \infty$) for $\gamma \leq 3$, leading to a vanishing epidemic threshold, i.e. any disease can invade the system, whatever its infection rate^{14,15}. For $\gamma > 3$, on the other hand, HMF predicts a finite threshold. This result has usually been interpreted in terms of the leading role of the hubs (the vertices with largest degree in the network) as the elements sustaining the

epidemic activity in the network, whenever they have a sufficiently large degree to make the second moment $\langle q^2 \rangle$ diverge (i.e. when $\gamma \leq 3$)⁹.

More refined approaches than HMF, incorporating the effects of the quenched topological structure of the network, but still neglecting dynamical correlations^{16,17,18}, predict that the epidemic threshold for SIS is in general set by the largest eigenvalue Λ_N of the adjacency matrix, i.e. $\lambda_c^E = 1/\Lambda_N$. This finding, combined with the scaling of Λ_N (computed by Chung et al. for a class of finite graphs with degrees distributed according to a power-law¹⁹), $\Lambda_N \sim \max[\sqrt{q_{max}}, \langle q^2 \rangle / \langle q \rangle]$, leads to a threshold²⁰

$$\lambda_c^E \simeq \begin{cases} 1/\sqrt{q_{max}} & \gamma > 5/2 \\ \frac{\langle q \rangle}{\langle q^2 \rangle} & 2 < \gamma < 5/2 \end{cases} . \quad (1)$$

Equation (1) implies that the epidemic threshold vanishes in the thermodynamic limit in power-law distributed networks for any value of γ , even larger than 3, as long as q_{max} is a growing function of the network size N , in agreement with previous results for SIS^{21,22,23}. In this perspective, it would be the hub, or most connected vertex, the main responsible of maintaining the epidemic activity and correspondingly setting the threshold²⁰.

The relevance of hubs has been however recently called into question by Kitsak et al.²⁴, who pointed out that in some real networks, the most efficient spreaders are located at the innermost, dense core of the network, as identified by means of a k -core decomposition²⁵ (see Methods and Figure 1). In this alternative view, it is thus the nucleus of high k -core index which sustains epidemic activity, independently of the degree of the vertices it is composed of.

Inspired by these results, here we analyze in detail the role of the hub and of the core of the network for the onset of epidemic spreading on complex topologies. By means of theoretical arguments and extensive numerical simulations, we are able to show that the leading mechanism governing the dynamics depends on the network features, in particular on the strength of the degree heterogeneity, as measured by the degree exponent γ . The analysis of real networks allows to determine additionally the critical role of degree correlations in suppressing or enhancing the relevant mechanism. The findings presented in this work represent an advancement in the understanding of the underlying mechanisms that control

the behavior of epidemic processes on complex heterogeneous networks. By identifying with precision the set of vertices ultimately responsible for the epidemic activation, our results open the path for the formulation of immunization strategies^{1,26} specifically tailored for each particular network configuration considered. Moreover, our results can find application in other, more general spreading processes, such as rumor, behavior or information spreading in networks^{27,28}, as well as other dynamical processes ruled by the largest eigenvalue of the adjacency matrix, such as synchronization phenomena²⁹.

Results

Activation mechanisms for SIS in uncorrelated networks

In the case of the SIS model the expression of the threshold λ_c for $\gamma > 5/2$ can be understood by considering the largest hub and its neighbors as a star network of size $q_{max} + 1$. Such a system has, in isolation, a threshold $\lambda_c^{star} = 1/\sqrt{q_{max}}$ and is thus capable, all by itself and independently of the degree of the rest of the vertices, to propagate the infection to a finite fraction of the network, leading to a stable endemic state whenever $\lambda > \lambda_c^{star}$ ²⁰. It is therefore the most connected vertex which singlehandedly can keep the epidemic activity alive, setting in this way the global threshold for activity in the system. The change for $\gamma < 5/2$ in equation (1) is however surprising and hints towards the possibility of different activation mechanisms for different γ values, thus challenging the belief in the preponderant role of hubs, which has become common wisdom in network science⁹. The results of Kitsak *et al.* would fit in place, pointing towards a preponderant role for the innermost core of the network. However, while the picture presented by Kitsak *et al.*²⁴ is compelling for the SIR model, the case of the SIS deserves a closer look.

In order to shed light on this issue, we have performed extensive numerical simulations of the SIS process on synthetic uncorrelated scale-free networks with degree distribution $P(q) \sim q^{-\gamma}$, generated via the uncorrelated configuration model (UCM)³⁰ (see Supplementary Information online for more details). We have computed the density of infected vertices

in the whole network, and the same density when the dynamics takes place (in isolation) on the k -core of highest index (maximum k -core) and on the star-graph centered around the hub of the network, with degree q_{max} .

In Figure 2 we show the evolution of the recorded densities as a function of time for different values of the spreading rate λ , in networks with large and small degree exponents, namely $\gamma = 2.75$ and $\gamma = 2.1$. Our results show a remarkable dependence on the degree exponent: For large γ , the onset of a global stationary state takes place for the same values of λ for which the star-graph centered around the hub starts to be active, while the maximum k -core remains subcritical, with exponentially decaying activity. This behavior is a proof of the leading role of the hub as the main activation mechanism for large γ . For small values of γ , instead, the picture is opposite: For values of λ corresponding to a globally active network, the maximum k -core is in an active state, while the star-graph centered around the hub is inactive, indicating that now the maximum k -core is the trigger activating the whole system. Two observations are in order: The maximum k -core is the heart of the nucleus of most densely connected vertices in the network but it does not sharply coincide with it. Other nodes, belonging to k -cores of index slightly smaller, are also densely connected. The transition in the whole network is influenced also by these other nodes and therefore only approximately coincides with the transition of the maximum k -core. This explains why in Fig. 2 the whole network is fully active for $\lambda = 0.01$, while the maximum k -core is still around the transition. The second observation is that in uncorrelated networks the hub usually belongs to the maximum k -core. Yet the two activation mechanisms for epidemics are clearly distinct. In one case (hub triggered activation) the hub alone is able to sustain activity in the set of its neighbors and then propagate it to the rest of the system. In the other (maximum k -core triggered activation) the hub alone is not able to sustain activity: Only the presence of all densely connected vertices in the k -core allows them to collectively turn into the active state and propagate to the rest of the system.

This change of behavior with the degree exponent, which we have confirmed for different values of γ above and below $5/2$ (see Supplementary Information Figure S1 online), can be made more physically transparent by linking it analytically with the different thresholds

in equation (1). To do so, we estimate the threshold associated to the active maximum k -core, whose index is denoted as k_S . The maximum k -core has a degree distribution which is bounded and narrow, with minimum (k_S), average, and maximum degree scaling with size in the same way (see Supplementary Information Figure S2 online). Hence its epidemic threshold is well approximated by $\lambda_c^K = 1/\langle q \rangle \sim 1/k_S$. On the other hand, in Ref.³¹ the maximum k -core index k_S was determined as a function of the network topology, yielding for scale-free networks with $2 < \gamma < 3$

$$k_S \approx (\gamma - 2)(3 - \gamma)^{(3-\gamma)/(\gamma-2)} q_{max} \left(\frac{q_{min}}{q_{max}} \right)^{(\gamma-2)}, \quad (2)$$

where q_{min} is the minimum degree. Introducing this result into the formula for λ_c^K we obtain $\lambda_c^K \sim q_{max}^{\gamma-3}$. It is most noteworthy that the scaling behavior of the maximum k -core threshold λ_c^K takes the exact same form as the eigenvalue threshold for $\gamma < 5/2$ in equation (1). This observation provides a physical interpretation of the different activation mechanisms and associated thresholds in uncorrelated scale-free networks: When $\gamma < 5/2$, the epidemic transition is collectively triggered by the vertices in the innermost core and the threshold is correspondingly given by $\lambda_c \sim 1/\langle q^2 \rangle \sim q_{max}^{\gamma-3}$, as in HMF theory. On the other hand, for $\gamma > 5/2$, the hub triggers the global activity, and the threshold is given by $\lambda_c \simeq 1/\sqrt{q_{max}}$. An additional inspection of the numerical values of the different thresholds (see Supplementary Information Table S1 online) shows that the thresholds computed from the numerical estimation of the largest eigenvalue of the adjacency matrix are in very good agreement with the predictions of Eq. (1), perfectly accounting for the results observed in Figure 2.

The SIR model

We turn now our attention to the SIR model, which, contrarily to the steady-state dynamics of the SIS model, exhibits transient outbreaks characterized by the number of infected individuals, totaling a finite fraction of the system only above the epidemic threshold. Evidence that the hub plays no special role in SIR dynamics comes from considering this process on a star network of size $q_{max} + 1$. For an epidemics starting from a randomly chosen vertex,

the average final density of infected nodes takes the form (see Methods)

$$R = \lambda^2 + \frac{1 + 2\lambda}{q_{max}}. \quad (3)$$

Hence the threshold in a star network, defined by the value of λ above which R takes a given fixed finite value, is a constant independent of q_{max} , in the limit of large q_{max} . The hub cannot therefore be the ultimate trigger of global outbreaks in the SIR model, and this role must instead be played by the maximum k -core for any value of γ , in accordance with Ref.²⁴. Additional support for this view comes from extending to the SIR case the maximum k -core threshold argument presented for the SIS model. Approximating again the maximum k -core as a narrowly distributed graph of average degree $\langle q \rangle \sim k_S$, a threshold is obtained from MF theory of the form $\lambda_c^{SIR,K} \sim 1/k_S$. Given the form of k_S in equation (2), this threshold scales in the large network limit in exactly the same form as the HMF prediction, namely $\lambda_c^{SIR} = \langle k \rangle / [\langle k^2 \rangle - \langle k \rangle]$ ^{32,33}. The conclusion is that in the SIR model it is always the maximum k -core which controls epidemic spreading and sets the threshold to the HMF value. This picture is substantiated in Figure 3, where we consider the SIR model on UCM networks with different values of γ , keeping track, as a function of λ , of the density of infected individuals in the global network, in the maximum k -core and in the star-graph centered around the hub. As we can see, the position of the transition to a finite fraction of infected vertices is closely correlated in the whole network and the maximum k -core, while the behavior of the hub conforms to the prediction of equation (3).

Effects of correlations

The scenario discussed so far applies to the case of uncorrelated networks, where the probability that a random edge is connected to a vertex of degree q is proportional to $qP(q)$ ⁵. Real networks, however, present in most cases some level of degree correlations¹⁰, as measured by the Pearson coefficient³⁴ or by the average degree of the nearest neighbors (ANN) of the vertices of given degree, $\bar{q}_{nn}(q)$ ³⁵. In order to ascertain their effect on the relevant epidemic mechanisms, we have considered the SIS process on several instances of correlated real networks (see Methods): An Internet map at the autonomous system (AS) level, the

social network of pretty-good-privacy (PGP), and the network of actors co-starring in Hollywood movies (Movies). All these networks have a degree distribution compatible with a power law, with an exponent close to 2 (see Supplementary Information Figure S3 online) and a range of degree correlations (see Supplementary Information Figure S4 online). In this case, according to our arguments above and neglecting correlations, one would expect the transitions to be ruled by the corresponding maximum k -cores. This fact is confirmed for the Movies and PGP networks, by the SIS simulations presented in Figure 4, showing that the transition occurs simultaneously for the maximum k -core and the whole system, while the star-graph centered around the hub remains inactive. In the case of the AS network, instead, the picture is surprisingly the opposite, and it is apparently the hub the responsible of the epidemic transition, see Figure 4. The situation is still more complex when one considers other, larger AS maps (see Supplementary Information). In fact, as it turns out from the analysis of numerical simulations (see Supplementary Information Figure S5 online), the general situation in AS maps is that we can reach in the network an active, infected state for values of λ for which both the hub and the maximum k -core are apparently subcritical. This observation hints towards a mixing of activation mechanisms for the particular case of AS networks. This discrepancy between the AS and the other networks, confirmed by the inspection of the values of the different thresholds (see Supplementary Information Table S2 online) can be attributed to the presence of strong degree correlations. Measuring them by means of the auxiliary ANN function \bar{q}_{nn} , we can observe that the AS network is strongly correlated, with \bar{q}_{nn} decaying as $q^{-0.5}$. Moreover, these correlations are so strong that they do not wash away even after randomizing the network, as they do in the PGP and Movies networks (see Supplementary Information Figure S4 online). Strong disassortative correlations reduce the interconnections of vertices of high-degree, suppressing in this way the index k_S of the maximum k -core and reducing the number of vertices that compose it. This situation, i.e. a very large hub coupled with a relatively small maximum k -core, leads to a mixing of both mechanisms that does not allow to make explicit prediction about the most relevant one. Similar or opposite effects (i.e. strong assortativity enhancing the role of the maximum k -core) can be found in networks generated by means of the Weber-Porto algorithm³⁶, a

modification of the configuration model that generates graphs with prescribed degree distribution and correlations of tunable strength (see Supplementary Information Figure S6 and Table S2 online).

Discussion

The rationalization of the different mechanisms that keep an epidemic process alive in a heterogeneous substrate turns out to be a more complex issue than previously believed. In fact, two different subsets of vertices (either the hub or the innermost core of the network) can take the role of “super-spreaders” of the infection, depending on the nature of the epidemic process and on the topological features of the underlying network. In processes with no steady state, such as the SIR model, the innermost core is the main trigger activating infection and setting the value of the epidemic threshold. On the contrary, in processes allowing an endemic steady-state, the actual activation mechanism depends essentially on the degree of heterogeneity of the network. This simple picture, valid for uncorrelated networks, can be however modified by the presence of strong degree correlations, which can shift the weight towards one or the other mechanism. These observations call for further theoretical research on epidemic processes on strongly correlated networks. On the other hand, our results might find practical applications in the implementation of optimized immunization strategies, which can be designed to target the actual “super-spreaders” of a given contact network. Moreover, our work could turn out to be relevant to other types of spreading processes, such as information, behavior or rumor spreading, and more in general, to dynamical processes whose behavior and possible transitions are ruled by the value of the largest eigenvalue of the adjacency matrix, such as, for example, synchronization phenomena. Also, for these latter dynamics, our results call for additional investigations on the existence and interplay of the relevant activation mechanisms.

Methods

The SIS and SIR epidemic models

In the SIS model, individuals can be in one of two states, either susceptible or infected. Susceptibles become infected by contact with infected individuals, with a rate equal to the number of infected contacts times a given spreading rate λ . Infected individuals on the other hand become healthy again with a rate μ that can be taken arbitrarily equal to unity, thus setting the characteristic time scale. This model allows thus individuals to contract the infection time and again, making possible, in the infinite population limit, a sustained infected steady state (endemic state). This occurs for values of λ larger than the epidemic threshold λ_c , while for $\lambda < \lambda_c$ the epidemics lasts only for a finite time and asymptotically all individuals are healthy.

In the SIR model, on the other hand, individuals can be in one of three different states: susceptible, infected and recovered (or removed). The dynamical rule for susceptible individuals is the same as for SIS. With a rate μ (again set to unity) infected individuals change their state and recover. Recovered individuals are completely inert and cannot become infected again. With this dynamics the system always reaches asymptotically an absorbing state with only susceptible or removed individuals and no infected ones. A threshold λ_c separates a regime where outbreaks reach a finite fraction of the individuals (i.e. the final density of removed individuals is finite) from a regime $\lambda < \lambda_c$ where only an infinitesimal fraction of individuals is hit.

The k -core decomposition.

The k -core decomposition is an iterative procedure to classify vertices of a network in layers of increasing density of connections. Starting with the full graph one removes the vertices with only one connection (degree $q = 1$). This procedure is then repeated until only nodes with degree $q \geq 2$ are left. The removed nodes constitute the $k = 1$ -shell and those remaining are the $k = 2$ -core. At the next step all vertices with degree $q = 2$ removed, thus leaving

the $k = 3$ -core. The procedure is repeated iteratively. The maximum k -core (of index k_S) is the set of vertices such that one more iteration of the procedure removes all of them. Notice that all vertices of the k -core of index k have degree larger than or equal to k . Figure 1 shows an example of the k -core decomposition performed on a small network of size $N = 30$ and largest degree $q_{max} = 10$.

SIR model on a star-graph

Let us consider the SIR process on a star network of size $q_{max} + 1$. The process starts from a randomly infected vertex, and proceeds till all infected vertices become eventually removed. We want to compute the average final density of removed vertices at the end of an outbreak, which is given by $R = r/(q_{max} + 1)$, where r is the total number of removed vertices. Let us define $r = 1 + r^*$, where r^* is the number of removed vertices from secondary infections. The outcome of the process will depend on whether the initial infected site is the hub or a leaf (a vertex of degree 1). If the infection starts in the hub, which will happen with probability $p_h = 1/(q_{max} + 1)$, the average number of secondary infected vertices will be $\langle r^* \rangle_h = \lambda q_{max}$. On the contrary, the infection can start in a leaf with probability $p_l = q_{max}/(q_{max} + 1)$. In this case, the hub can become infected with probability λ , and from there, spread the infection to the remaining $q_{max} - 1$ susceptible leaves. Therefore, in this case the average number of secondary infections is $\langle r^* \rangle_l = \lambda[1 + \lambda(q_{max} - 1)]$. The average total number of removed sites at the end of the spreading process will therefore be

$$\begin{aligned} \langle r \rangle &= 1 + \langle r^* \rangle = 1 + p_h \langle r^* \rangle_h + p_l \langle r^* \rangle_l = 1 + \frac{\lambda q_{max}}{q_{max} + 1} + \frac{q_{max}}{q_{max} + 1} \lambda [1 + \lambda(q_{max} - 1)] \\ &= 1 + \frac{\lambda q_{max}}{q_{max} + 1} [2 + \lambda(q_{max} - 1)] \simeq 1 + 2\lambda + \lambda^2 q_{max}, \end{aligned}$$

where the last expression is valid in the limit of large q_{max} . From here it follows the expression for the average final density of infected vertices, equation (3), valid also in the limit of large q_{max} .

Quantitative features of the real networks considered

We consider in our analysis the following three real networks datasets:

Internet map at the Autonomous System level (AS). Map of the Internet collected at the Oregon route server. Vertices represent autonomous systems (aggregations of Internet routers under the same administrative policy), while edges represent the existence of border gateway protocol (BGP) peer connections between the corresponding autonomous systems³⁷.

Pretty-good-privacy network (PGP). Social network defined by the users of the pretty-good-privacy (PGP) encryption algorithm for secure information exchange. Vertices represent users of the PGP algorithm. An edge between two vertices indicates that each user has signed the encryption key of the other³⁸.

Actor collaboration network (Movies). Network of movie actor collaboration obtained from the Internet Movie Database (IMDB). Each vertex represents a movie actor. Two actors are joined by an edge if they have co-starred at least one movie³⁹.

The relevant topological features of the different maps are summarized in Table 1.

Acknowledgments

RP-S acknowledges financial support from the Spanish MICINN, under project No. FIS2010-21781-C02-01, and the Junta de Andalucía, under project No. P09-FQM4682, as well as additional support through ICREA Academia, funded by the Generalitat de Catalunya.

References

1. Cohen, R., Havlin, S. & ben Avraham, D. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **91**, 247901 (2003).
2. Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. *ACM Trans. Web* **1** May (2007).
3. Daley, D. J. & Kendall, D. G. Epidemics and rumours. *Nature* **204**, 1118 (1964).

4. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
5. Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks. *Advances in Physics* **51**, 1079–1187 (2002).
6. Newman, M. The Structure and Function of Complex Networks. *SIAM Review* **45**, 167 (2003).
7. Keeling, M. J. & Eames, K. T. D. Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307 (2005).
8. Anderson, R. M. & May, R. M., *Infectious diseases in humans*, (Oxford University Press, Oxford, 1992).
9. Barrat, A., Barthélemy, M. & Vespignani, A., *Dynamical Processes on Complex Networks*, (Cambridge University Press, Cambridge, 2008).
10. Serrano, M. A., Boguñá, M., Pastor-Satorras, R. & Vespignani, A. Correlations in complex networks. In *Large scale structure and dynamics of complex networks: From information technology to finance and natural sciences*, Caldarelli, G. & Vespignani, A., editors, 35–66 (World Scientific, Singapore, 2007).
11. Boguñá, M., Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in complex networks with degree correlations. In *Statistical Mechanics of Complex Networks*, Pastor-Satorras, R., Rubí, J. M. & Díaz-Guilera, A., editors, volume 625 of *Lecture Notes in Physics*. Springer Verlag, Berlin (2003).
12. Lloyd, A. L. & May, R. M. How viruses spread among computers and people. *Science* **292**, 1316–1317 (2001).
13. Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128 (2002).

14. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
15. Boguñá, M., Pastor-Satorras, R. & Vespignani, A. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.* **90**, 028701 (2003).
16. Wang, Y., Chakrabarti, D., Wang, C. & Faloutsos, C. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems (SRDS'03)*, 25–34. IEEE Computer Society, Los Alamitos, CA, USA (2003).
17. Gómez, S., Arenas, A., Borge-Holthoefer, J., Meloni, S. & Moreno, Y. Discrete-time markov chain approach to contact-based disease spreading in complex networks. *Europhys. Lett.* **89**(3), 38009 (2010).
18. Van Mieghem, P. The N-intertwined SIS epidemic network model. *Computing* **93**, 147—169 (2011).
19. Chung, F., Lu, L. & Vu, V. Spectra of random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **100**, 6313–6318 (2003).
20. Castellano, C. & Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701 (2010).
21. Ganesh, A., Massoulié, L. & Towsley, D. The effect of network topology on the spread of epidemics. In *IEEE INFOCOM*, 1455–1466, (2005).
22. Chatterjee, S. & Durrett, R. Contact processes on random graphs with power law degree distributions have critical value 0. *Annals of Probability* **37**, 2332–2356 (2009).
23. Durrett, R. Some features of the spread of epidemics and information on a random graph. *Proc. Natl. Acad. Sci. USA* **107**, 4491–4498 (2010).

24. Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E. & Makse, H. A. Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
25. Seidman, S. B. Network structure and minimum degree. *Social Networks* **5**, 269 –287 (1983).
26. Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65**(3), 036104 Feb (2002).
27. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**(5996), 1194–1197 09 (2010).
28. Lü, L., Chen, D.-B. & Zhou, T. The small world yields the most effective information spreading. *New Journal of Physics* **13**(12), 123005 (2011).
29. Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y. & Zhou, C. Synchronization in complex networks. *Physics Reports* **469**(3), 93 – 153 (2008).
30. Catanzaro, M., Boguñá, M. & Pastor-Satorras, R. Generation of uncorrelated random scale-free networks. *Phys. Rev. E* **71**, 027103 (2005).
31. Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. k -core organization of complex networks. *Phys. Rev. Lett.* **96**, 040601 (2006).
32. Cohen, R., Erez, K., ben Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.* **85**(21), 4626–4628 Nov (2000).
33. Newman, M. E. J. The spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128 (2002).
34. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
35. Pastor-Satorras, R., Vázquez, A. & Vespignani, A. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* **87**, 258701 (2001).

- 36. Weber, S. & Porto, M. Generation of arbitrarily two-point-correlated random networks. *Phys. Rev. E* **76**, 046111 (2007).
- 37. Pastor-Satorras, R. & Vespignani, A., *Evolution and structure of the Internet: A statistical physics approach*, (Cambridge University Press, Cambridge, 2004).
- 38. Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A. & Arenas, A. Models of social networks based on social distance attachment. *Phys. Rev. E* **70**, 056122 (2004).
- 39. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

Tables

	N	$\langle q \rangle$	q_{max}	k_S
Movies	81860	89.532	3789	359
PGP	10680	4.554	205	31
AS	11174	4.190	2389	17

Table 1: Topological features of the real network datasets considered. Network size N , average degree $\langle q \rangle$, degree of the largest hub q_{max} , and index of the maximum k -core, k_S .

Figure Captions

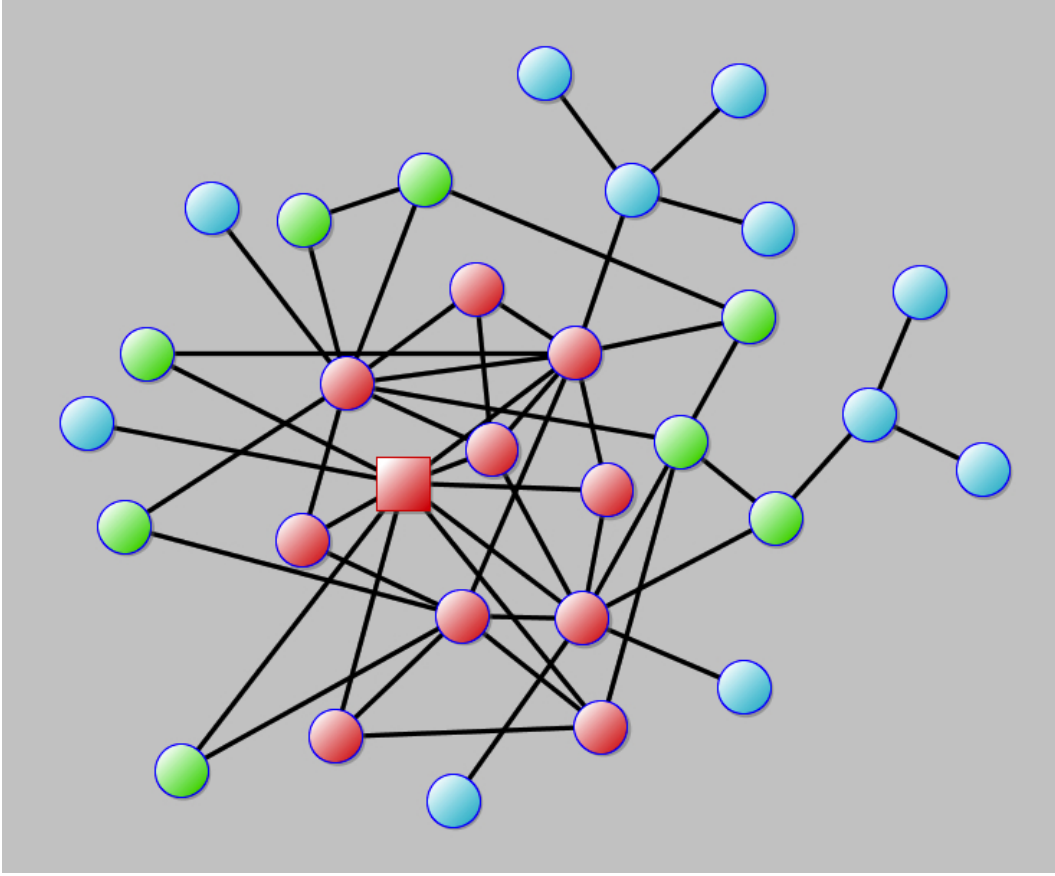


Figure 1: Visual representation of the k -core decomposition of a small network of size $N = 30$ and maximum degree $q_{max} = 10$. Blue vertices belong to the $k = 1$ shell and green vertices to the $k = 2$ shell. The maximum k -core, with $k_S = 3$, is composed by the red vertices. The hub (vertex with largest degree) is represented as a square.

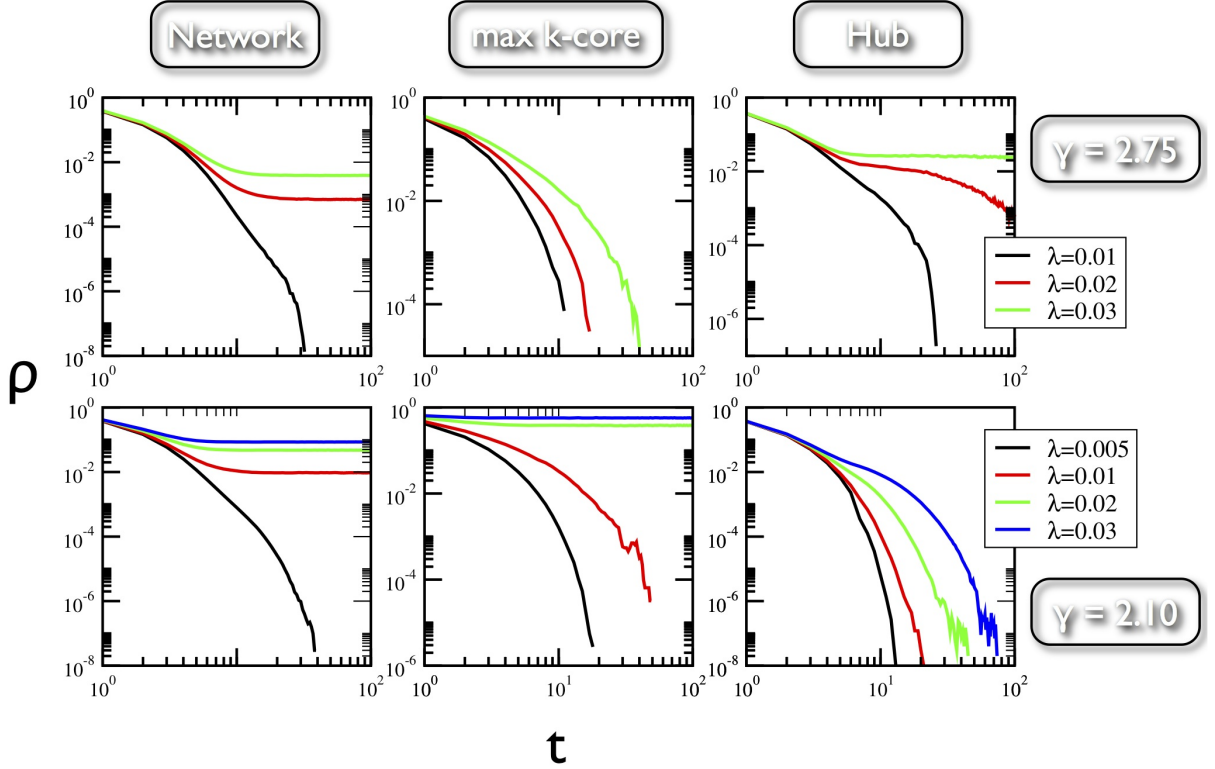


Figure 2: Average density of infected vertices as a function of time, $\rho(t)$, in the SIS model on uncorrelated scale-free networks generated by means of the UCM algorithm. We consider networks with widely separated degree exponent and different size, namely $\gamma = 2.75$, $N = 3 \times 10^7$ and $\gamma = 2.1$, $N = 10^6$. The different columns correspond to the average density computed when the dynamics runs over the whole network (left), only over the maximum k -core of the network (center), and only over the largest hub (right), considered as an isolated star network. The different colors correspond to different values of the spreading rate λ . For small $\gamma = 2.1$ (bottom row), the onset of the global steady state is correlated with the active state of the epidemics on the maximum k -core, while it corresponds to a subcritical state for the hub. This observation indicates that in this case the maximum k -core is responsible for the overall activity in the network. For large $\gamma = 2.75$, on the other hand, the global active state is linked to an active hub and a subcritical maximum k -core, signaling that it is the former mechanism the one keeping activity on a global scale.

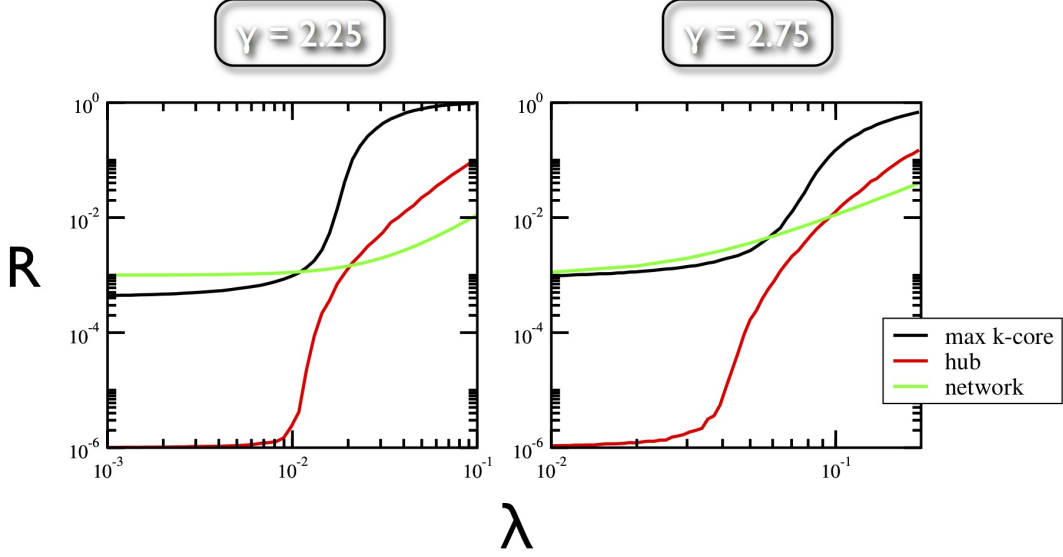


Figure 3: Total density R of infected vertices, as a function of the spreading rate λ , computed after an epidemic outbreak in the SIR model on uncorrelated scale-free networks generated by means of the UCM algorithm. The degree exponents considered are $\gamma = 2.25$ (left) and $\gamma = 2.75$ (right), with network sizes $N = 10^6$. The line colors correspond to the SIR dynamics restricted to the maximum k -core (black), to the largest hub (green), and on the whole network (red). The value of λ after which a macroscopic fraction of the network becomes infected is correlated in the whole network and the maximum k -core, while the infection pattern on the hub conforms with the theoretical expression in equation (3). These results indicate the crucial role of the maximum k -core in keeping the SIR activity on a large scale in networks, independently of the behavior of the hub.

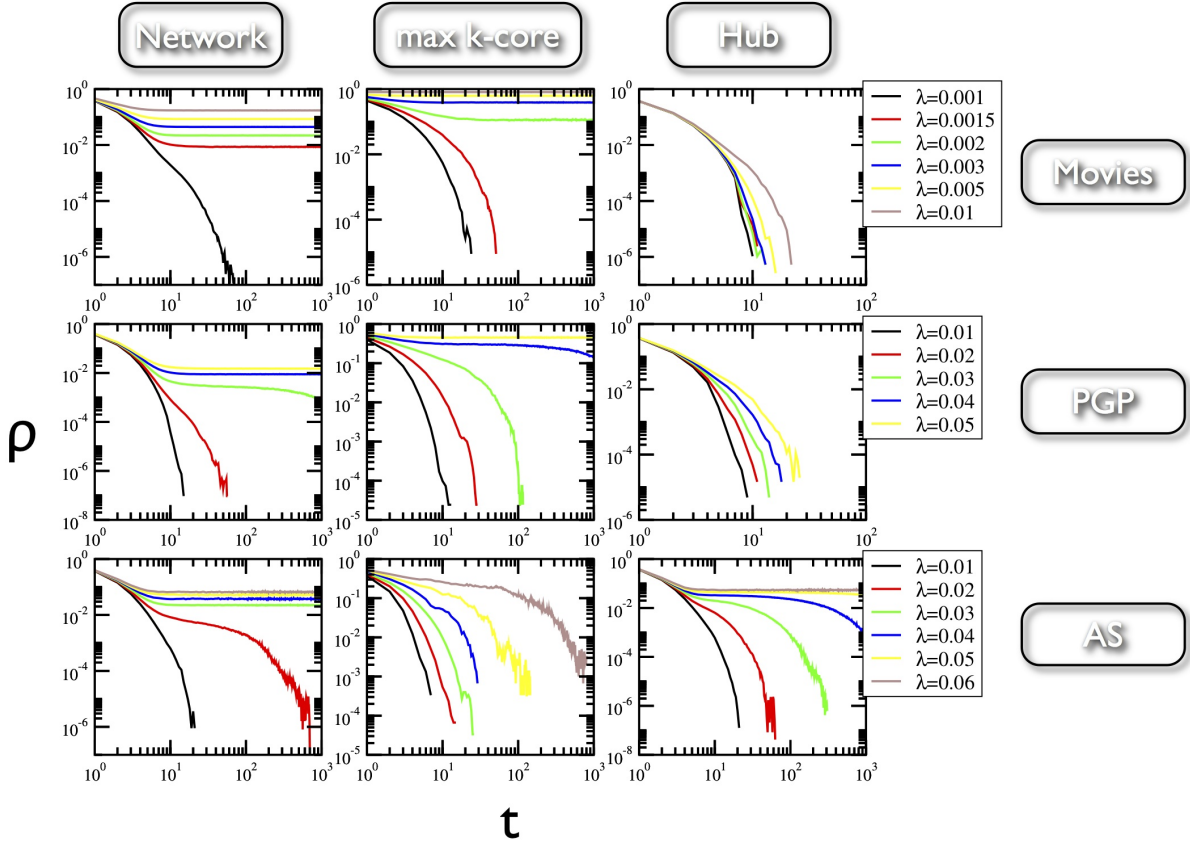


Figure 4: Average density of infected vertices as a function of time, $\rho(t)$, for the SIS model on three instances of real correlated networks: The network of actors co-starring in Hollywood movies (Movies), the social network of pretty-good-privacy (PGP), and an Internet map at the autonomous system level (AS); see Methods for further details of the networks. As in Fig. 2, columns refer to the activity on the whole network (left) and restricted to the maximum k -core (center) or the largest hub (right). Line colors indicate different values of the spreading rate considered in the networks. In the Movies and PGP networks, the maximum k -core dominates the transition, as expected due to the small degree exponent of all three networks ($\gamma \sim 2$). Surprisingly, in the AS network it is the activation of the hub the dominant mechanism setting in the steady state. This different behavior must be attributed to the very strong correlations present in the AS map (see main text).

SUPPLEMENTARY MATERIAL

1 Details of SIS numerical simulations.

The numerical simulations of SIS dynamics have been performed in the standard rejection-free way¹: At each time step one node from the list of infected vertices is chosen at random and removed; then, each of its neighbors is considered and, if healthy, infected with probability λ . Time is incremented by the inverse of the total number of occupied nodes. Substrate networks are created using the uncorrelated configuration model (UCM)².

Initially, all nodes are taken to be infected. No variation is expected for any finite fraction of infected nodes in the initial state. Spreading experiments, with a single infected vertex in the initial state, are an interesting topic for future work.

Results are averaged over many realizations (at least 10) of the dynamical process on a single network. We have not averaged over different network samples, because fluctuations in the various geometrical quantities (max k-core order, max k-core size, hub connectivity) make averaged results less clear (see Ref.³).

The system sizes considered for larger γ are larger than for smaller values of γ . This is motivated by the need to increase the separation between $1/\sqrt{q_{max}}$ and $\langle q \rangle / \langle q^2 \rangle$, in order to allow for a clear distinction between the two mechanisms.

2 Results for additional values of γ .

In Fig. 5 we report the results of additional simulations of the SIS model on uncorrelated power-law distributed networks with exponent γ , for values of γ different from those reported in the main text. These results confirm that the activation mechanism triggering the epidemic transition is the largest hub or the innermost maximum k -core, depending on whether γ is larger or smaller than $5/2$.

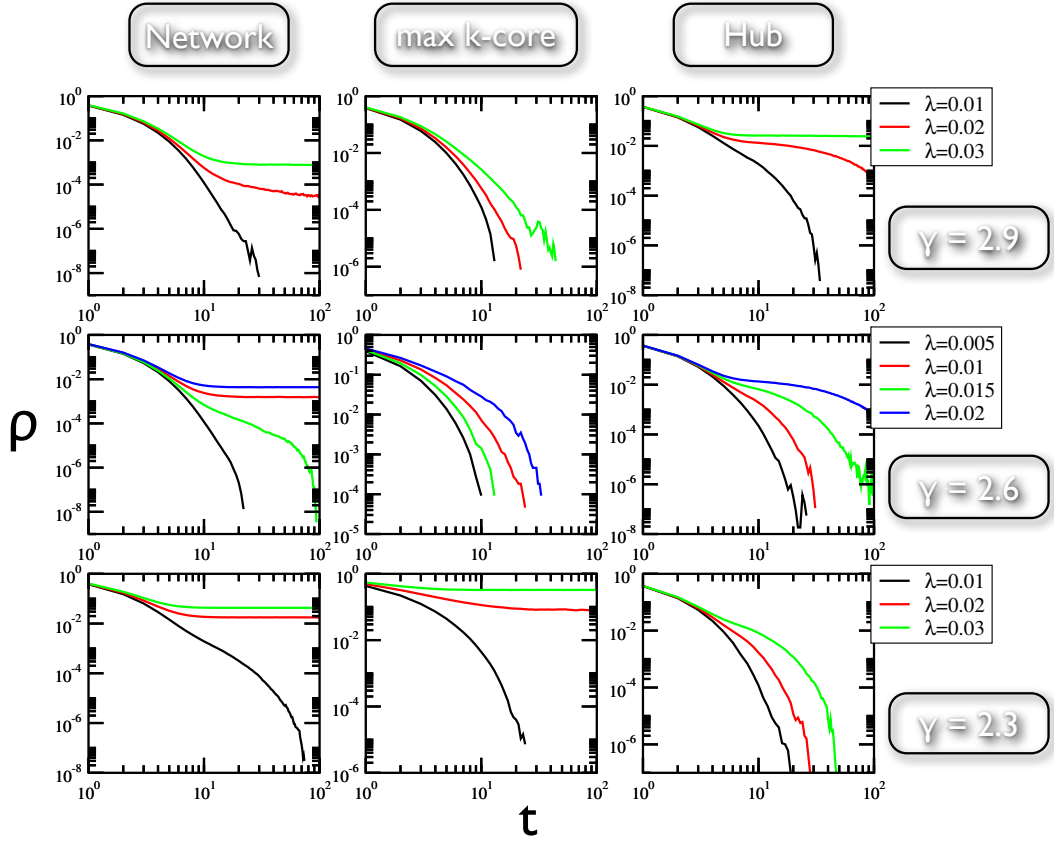


Figure 5: Average density of infected vertices as a function of time, $\rho(t)$, in the SIS model on uncorrelated scale-free networks generated by means of the UCM algorithm. We consider networks with $\gamma = 2.9$, $\gamma = 2.6$ ($N = 3 \times 10^7$) and $\gamma = 2.3$ ($N = 10^6$). The different columns correspond to the average density computed when the dynamics runs over the whole network (left), only over the maximum k -core of the network (center), and only over the largest hub (right), considered as an isolated star network. The different colors correspond to different values of the spreading rate λ . For $\gamma = 2.3$ (bottom row), the onset of the global steady state is correlated with the active state of the epidemics on the maximum k -core, while it corresponds to a subcritical state for the hub. This observation indicates that in this case the maximum k -core is responsible for the overall activity in the network. For large $\gamma > 5/2$, on the other hand, the global active state is linked to an active hub and a subcritical maximum k -core, signaling that it is the former mechanism the one keeping activity on a global scale.

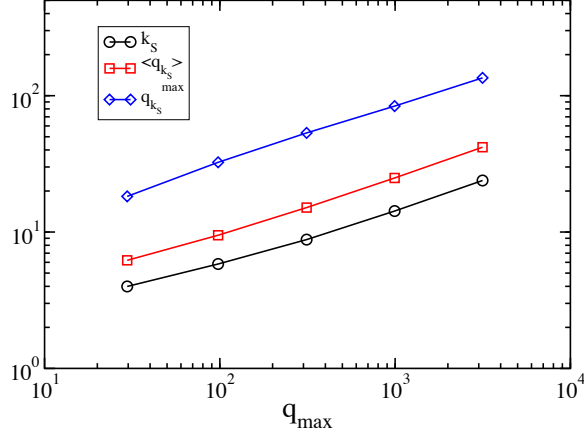


Figure 6: Topological properties of the of the maximum k -core as a function of network size in uncorrelated scale-free networks with degree exponent $\gamma = 2.5$.

3 Degree distribution of the maximum k -core.

We study the topological properties of the maximum k -core in uncorrelated scale-free networks generated using the UCM model. We focus in particular on the average degree $\langle q_{k_S} \rangle$ of the maximum k -core, and on its maximum $q_{k_S}^{max}$ and minimum $q_{k_S}^{min} = k_S$ degree. In Fig. 6 we plot this quantities averaged over networks with different size N as a function of the network's maximum degree $q_{max} \sim N^{1/2}$. The networks considered have a degree exponent $\gamma = 2.5$. From this figure, we conclude that all the computed properties of the maximum k -core scale in the same way with the network size (i.e. are proportional to each other), indicating that the degree distribution remains rather narrow for any network size. This observation justifies the assumption that heterogeneous mean-field theory determines correctly the scaling of the epidemic threshold on the maximum k -core.

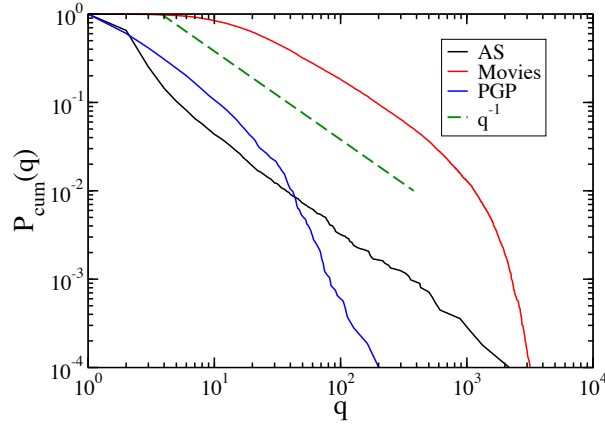


Figure 7: Cumulated degree distribution of the real network datasets considered. All plots show a power-law regime compatible with the form q^{-1} , indicating a degree distribution with a degree exponent $\gamma \sim 2$.

4 Degree distributions and correlations in real networks.

In Fig. 7 we report the degree distributions of the three real networks considered. In all cases the distributions decay slowly for large q with an exponent close to $\gamma = 2$.

In Fig. 8(a) we plot on the other hand the average degree of the nearest neighbors (ANN) of a node of degree q for the three real networks considered⁴. The Movies and PGP networks exhibit a mild assortative structure⁵: vertices with high degree tend to be connected in average with vertices of high degree, while low degree vertices are connected to each other. On the contrary, the AS network is strongly disassortative, i.e. vertices with large q are typically connected to nodes with small q , and viceversa. The different strength of degree correlations is highlighted by computing the $\bar{q}_{nn}(q)$ function for the same networks after a randomizing (degree-preserving) procedure is applied⁶: two edges are randomly selected and two vertices at their respective ends are swapped.

While the rewiring procedure completely destroys the weak assortative correlations of the Movies and PGP networks, Fig. 8(b), no effect can be spotted for the AS, indicating the

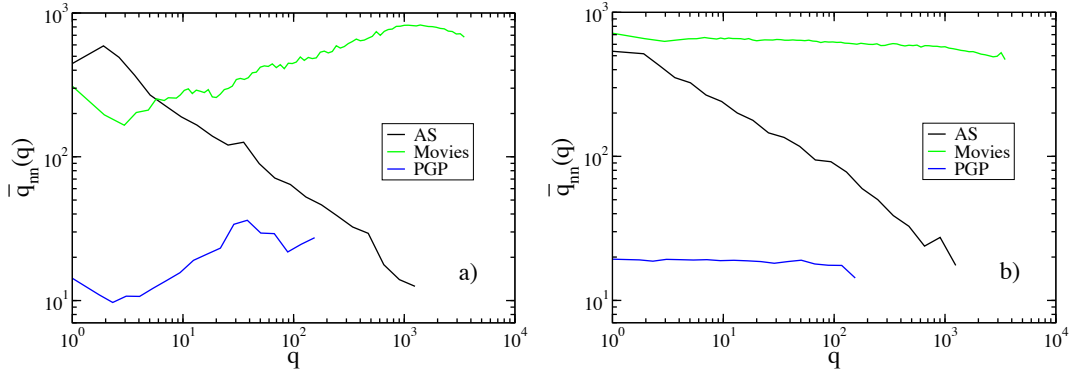


Figure 8: Average degree of the nearest neighbors (ANN) function, $\bar{q}_{nn}(q)$, for the three real networks considered. a) Original networks. b) Randomized networks, under application of the algorithm in Ref.⁶.

presence of very strong and robust correlations.

5 Values of the thresholds

The numerical values of the analytical estimates of the different thresholds discussed for the SIS in networks are reported in Table 2. We focus in particular on the networks considered in our manuscript, namely uncorrelated networks generated with the UCM algorithm and three real network datasets. The first column shows the values of $\lambda_c^E = 1/\Lambda_N$, the inverse of the largest eigenvalue of the adjacency matrix. In the second column, we find the value of $1/\sqrt{q_{max}}$, which is the prediction for uncorrelated networks with $\gamma > 5/2$. Finally, the third column presents the value of $\langle q \rangle / \langle q^2 \rangle$, which holds for $\gamma < 5/2$ and coincides with the HMF prediction. We observe that for UCM networks the predictions of Eq. (1) in the manuscript are very well obeyed, indicating that the estimates of Ref.⁷ are correct. With regards to real networks, the Movies network shows a threshold in good agreement with the prediction of Eq. (1) for $\gamma < 5.2$, i.e. close to the HMF prediction, indicating the relevant role of the maximum k -core. For the PGP network, the agreement with Eq. (1) is not precise, yet the threshold is closer to the HMF prediction and therefore the maximum k -core

drives the transition, as shown in Fig. 3 in the manuscript. The opposite is true for the AS network: the inverse of the largest eigenvalue is close to $1/\sqrt{q_{max}}$ and is much larger than the HMF prediction, despite a very slow decay of the degree distribution. The conclusion is that in this case the estimate of Ref.⁷ is largely incorrect, thus invalidating Eq. (1). This failure is originated by the disassortative correlations present in AS networks, induced by the presence of a very large hub, which mixes the role of the maximum k -core and the hub in the activation of the epidemics.

	$1/\Lambda_N$	$1/\sqrt{q_{max}}$	$\langle q \rangle / \langle q^2 \rangle$
UCM model			
$\gamma = 2.9, N = 3 \cdot 10^7$	0.01352	0.01356	0.03896
$\gamma = 2.75, N = 3 \cdot 10^7$	0.01308	0.01353	0.02312
$\gamma = 2.6, N = 3 \cdot 10^7$	0.01031	0.01351	0.01294
$\gamma = 2.3, N = 10^6$	0.01160	0.03168	0.01209
$\gamma = 2.1, N = 10^6$	0.00735	0.03163	0.00745
Real networks			
Movies	0.001223	0.01624	0.00168
PGP	0.02356	0.06984	0.05296
AS	0.016576	0.02045	0.003765

Table 2: Table of SIS thresholds for synthetic and real networks.

To further check the role of correlations and a large hub, we have considered an additional AS map, corresponding to later snapshot of the Internet, and thus characterized by a larger size. The characteristics of this second Internet map (AS2) are the following: $N = 24463$, $\langle q \rangle = 4.2799$, $k_S = 24$, $1/\Lambda_N = 0.01412$ $1/\sqrt{q_{max}} = 0.02012$ $\langle q \rangle / \langle q^2 \rangle = 0.003796$. In Figure 9 we present the results of numerical simulations of the SIS model performed on this AS2 map. We observe that, in this case, for values of λ corresponding to an active global network, neither the hub nor the maximum k -core are apparently in the active phase. This

result, which is also found in other AS maps we have considered, confirms that the presence of very strong correlations alters the clear picture observed in uncorrelated networks, leading to a mixing of the activation mechanisms.

6 SIS dynamics on Weber-Porto networks.

Recently, Weber and Porto⁸ have introduced a modified configuration model⁹ which allows to generate graphs with a given degree distribution and, simultaneously, a predetermined ANN function $\bar{q}_{nn}(q)$. We have performed simulations of the SIS dynamics on networks generated by the Weber-Porto method, and compared the density of active nodes as a function of λ when the dynamics takes place on the whole network or only on the maximum k -core or on a star-graph with $q_{max} + 1$ nodes, where q_{max} is the degree of the largest hub. Numerical values of the different theoretical thresholds, for different values of the degree exponent γ and of the exponent α of the ANN function, $\bar{q}_{nn}(q) \sim q^\alpha$, are reported in Table 3. Positive values of α imply assortative correlations, while $\alpha < 0$ is the signature of disassortative correlations. Results of numerical simulations are presented in Fig. 10.

	$1/\Lambda_N$	$1/\sqrt{q_{max}}$	$\langle q \rangle / \langle q^2 \rangle$
	Weber-Porto network model		
$\gamma = 2.1, \alpha = 0.5$	0.008548	0.05634	0.01798
$\gamma = 3.5, \alpha = -0.5$	0.064435	0.06482	0.14492
$\gamma = 2.1, \alpha = -0.5$	0.016137	0.01754	0.00353
$\gamma = 3.5, \alpha = 2$	0.021856	0.07001	0.14675

Table 3: Table of SIS thresholds for Weber-Porto networks.

In the first two cases presented in Table 3, correlations are expected to strengthen the mechanism already driving the transition for uncorrelated networks: indeed for $\gamma > 5/2$ disassortative correlations make the network more “star-like”, while for $\gamma < 5/2$ assortative

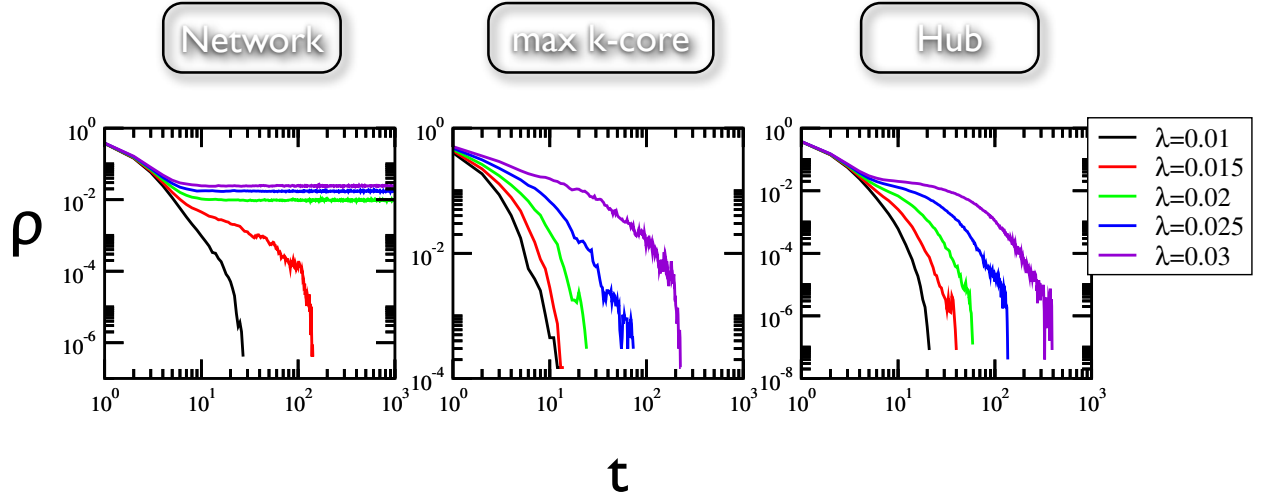


Figure 9: Average density of infected vertices as a function of time, $\rho(t)$, in the SIS model on the AS2 Internet map. The different columns correspond to the average density computed when the dynamics runs over the whole network (left), only over the maximum k -core of the network (center), and only over the largest hub (right), considered as an isolated star network. The different colors correspond to different values of the spreading rate λ .

correlations are expected to make the max k -core play an even more important role. This is confirmed by the data in Table 3:

- For $\gamma = 2.1$ and $\alpha = 0.5$, the threshold $1/\Lambda_N$ is smaller than the HMF prediction (which is in its turn smaller than $1/\sqrt{q_{max}}$).
- For $\gamma = 3.5$ and $\alpha = -0.5$ the threshold $1/\Lambda_N$ is in very good agreement with $1/\sqrt{q_{max}}$.

Eq. (1) in the manuscript turns out to be essentially correct. Numerical results in Supplementary Figure 10 confirm these findings. Notice that for $\gamma = 3.5$ and $\alpha = -0.5$ there is no k -core structure in the network. Numerical results (Fig. 10) confirm these findings.

In the other two cases, correlations act against the mechanism at work on uncorrelated networks. Are they able to completely perturb the picture and change the activation mechanism of the epidemics?

- For $\gamma = 2.1$ and disassortative networks ($\alpha = -0.5$), the threshold $1/\Lambda_N$ is larger than the HMF prediction and it coincides with $1/\sqrt{q_{max}}$. This indicates that the largest hub drives the global transition, at odds with the prediction of Eq. (1). Here correlations change completely the behavior, leading to a violation of the estimate of Λ_N on Ref.⁷. The direct numerical simulations, presented in Figure 10, confirm these findings.
- For $\gamma = 3.5$ and assortative correlations ($\alpha = 2$), the threshold $1/\Lambda_N$ is much smaller than $1/\sqrt{q_{max}}$ (which is in its turn smaller than the HMF prediction). This shows again that the result of Ref.⁷ does not hold. The direct numerical simulations, Figure 10, clearly shows that it is the maximum k -core which triggers the transition in this case.

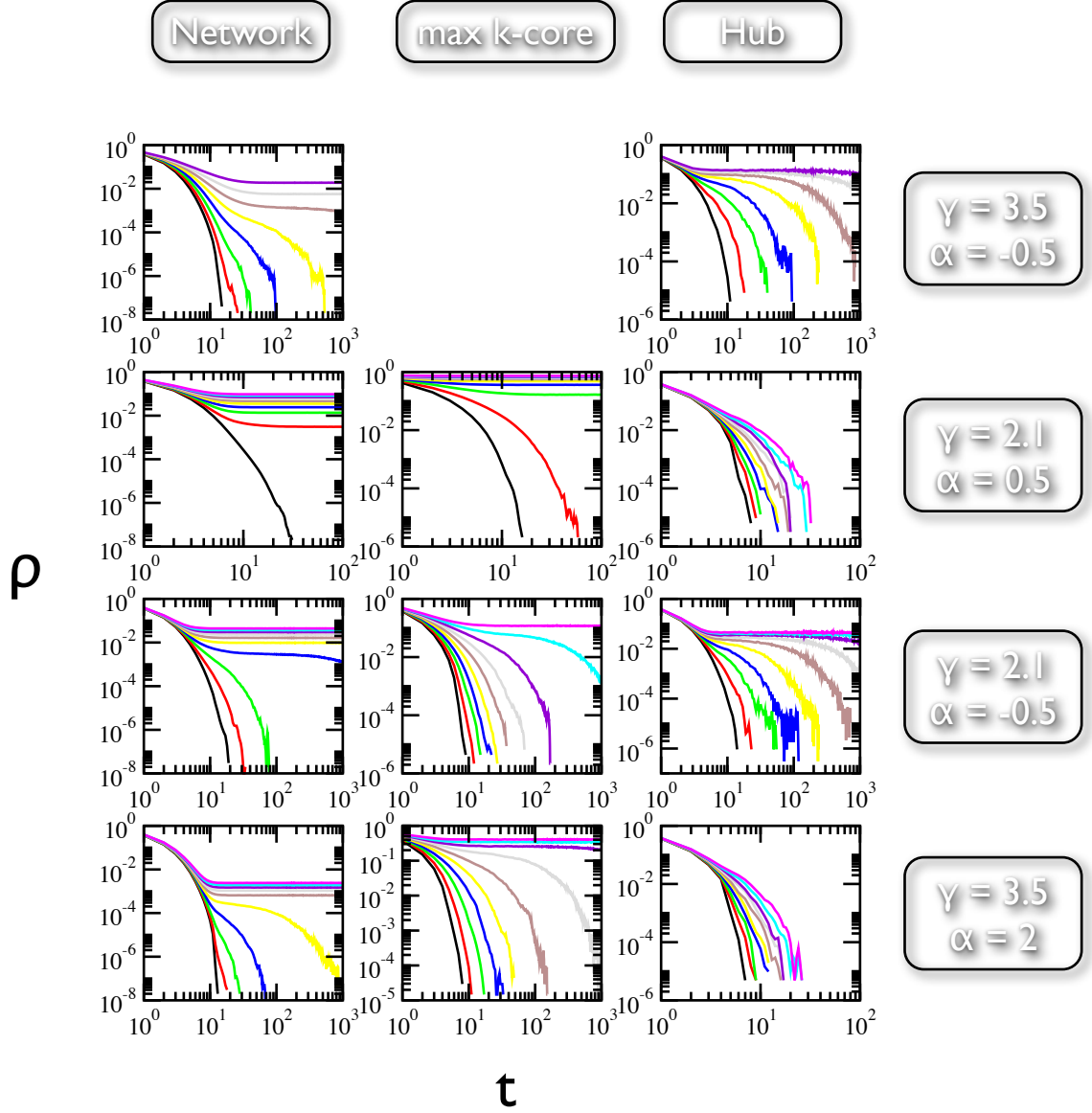


Figure 10: Density of infected vertices as a function of time in the SIS model on Weber-Porto networks with different values of γ and α . Network sizes $N = 10^4$. We plot the average density over the whole network (left column), over the maximum k -core (center column), and over the largest hub (right columns). The different colors correspond, from bottom to top of each plot, to increasing values of the spreading rate λ .

References

1. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
2. Catanzaro, M., Boguñá, M. & Pastor-Satorras, R. Generation of uncorrelated random scale-free networks. *Phys. Rev. E* **71**, 027103 (2005).
3. Castellano, C. & Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701 (2010).
4. Pastor-Satorras, R., Vázquez, A. & Vespignani, A. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* **87**, 258701 (2001).
5. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
6. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
7. Chung, F., Lu, L. & Vu, V. Spectra of random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **100**, 6313–6318 (2003).
8. Weber, S. & Porto, M. Generation of arbitrarily two-point-correlated random networks. *Phys. Rev. E* **76**, 046111 (2007).
9. Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* **6**, 161 (1995).